Mathematical Methods and Models in Biosciences
June 15–20, 2025, Sofia, Bulgaria
https://biomath.math.bas.bg/biomath/index.php/bmcs

Application of the information entropy and machine learning algorithms for the prediction of peptide binders to the HLA-DRB1*03:01 allele

Ivan Dimitrov, Mariyana Atanasova, Irini Doytchinova

Faculty of Pharmacy,
Medical University of Sofia, Bulgaria
idimitrov@pharmfac.mu-sofia.bg
matanasova@pharmfac.mu-sofia.bg
idoytchinova@pharmfac.mu-sofia.bg

The HLA-DRB1*03:01 allele is a known genetic risk factor for several autoimmune diseases, including systemic lupus erythematosus, type 1 diabetes, early-onset myasthenia gravis, Sjögren's syndrome, and autoimmune hepatitis type 1. As a key component of the antigen presentation pathway, HLA-DRB1*03:01 has also been associated with allergy and asthma. Predicting peptide binders to this allele could aid in identifying potential antigens linked to these critical autoimmune disorders.

In this study, we applied machine learning algorithms to a dataset of peptides with known binding affinity to the HLA-DRB1*03:01 allele. The information entropy for each peptide was calculated using numerical descriptors representing various physicochemical properties of the amino acids in the peptide sequence. This transformation converted peptide sequences into a numerical matrix, where each vector contained the calculated information entropy for different descriptors.

The dataset was split into training and test sets (80/20 ratio), and an iterative self-consistent algorithm was employed to determine the binding core of each peptide in the training set. Regression models were then used to predict binding affinity to the HLA-DRB1*03:01 allele. The derived models were validated on the test set, and an appropriate classification cutoff was determined.

The performance of the machine learning models was evaluated using classification metrics, revealing that the XGBoost and Support Vector Machine (SVM) models demonstrated superior results, achieving high sensitivity and accuracy.

*Keywords: information entropy, machine learning, HLA-DRB1*03:01 binding prediction*